UNIVERSITY OF PENNSYLVANIA

THE SUPERIORITY MEASURES OF THE PERFORMANCE OF FOURTH GRADE CHILDREN: AN ANALYSIS OF THE WITMER CLINICAL STANDARDS

THOMAS J. SNEE

A THESIS

IN PSYCHOLOGY

PRESENTED TO THE FACULTY OF THE GRADUATE SCHOOL IN
PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY

PHILADELPHIA
1933

THE SUPERIORITY MEASURES OF THE PERFORMANCE OF FOURTH GRADE CHILDREN: AN ANALYSIS OF THE WITMER CLINICAL STANDARDS

A THESIS

IN PSYCHOLOGY

PRESENTED TO THE FACULTY OF THE GRADUATE SCHOOL OF THE UNIVERSITY OF PENNSYLVANIA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

THOMAS J. SNEE

Reprint from
The Psychological Clinic, Vol. XXII, No. 1

PHILADELPHIA
1933

R. UNIV.
BIBLIOTHEEK
LEIDEN

Reprinted from The Psychological Clinic, Vol. XXII,
No. 1, March–May, 1933

THE SUPERIORITY MEASURES OF THE PERFORMANCE OF FOURTH GRADE CHILDREN: AN ANALYSIS OF THE WITMER CLINICAL STANDARDS

By Thomas J. Snee

University of Pennsylvania

Part I

The program of investigation inaugurated ten years ago by Dr. Lightner Witmer, Director of the Psychological Laboratory and Clinic of the University of Pennsylvania, with a view to obtaining results of a quantitative nature on a battery of tests by means of which the behavior of any child examined in the Psychological Clinic might be compared and rated with the performance of his group is nearing completion. Such norms are now at hand for the pre-school group,[1] for the first,[2] third,[3] fifth,[4] and sixth grades,[5] together with the fifteen year level,[6] while the second grade is now under investigation. The present study is concerned with the determination and analysis of the Witmer Clinical Standards for the Fourth Grade.

The general procedure followed in this study consisted of the examination of a group of children in the fourth grade after the method used in the previous investigations. The actual field work was begun in March, 1931, and completed in May, 1932.[7]

1 Hallowell, Dorothy. Mental Tests for Pre-School Children. *Psychol. Clin.*, 1928, 16, 233–276.

2 Easby-Grave, Charlotte. Tests and Norms at the Six Year Old Performance Level. *Psychol. Clin.*, 1924, 15, 261–300. *Cf.* Richards, T. W. Psychological Tests in the First Grade. *Psychol. Clin.*, 1932, 21, 235–242.

3 McDermott-Murphy, Genevieve. Unpublished manuscript in the Department of Psychology, University of Pennsylvania.

4 Murphy, Miles. The Ten Year Level of Competency. *Psychol. Clin.*, 1928, 17, 33–60.

5 Altmaier, Carl L. The Performance Level of Children in the Sixth Grade in Two Philadelphia Public Schools. *Psychol. Clin.*, 1931, 19, 233–257.

6 Leaming, Rebecca E. Tests and Norms for Vocational Guidance at the Fifteen Year Old Performance Level. *Psychol. Clin.*, 1922, 14, 193–220.

7 The author wishes to express his appreciation to Dr. Witmer, for the gift of his point of view and approach to the problems of clinical psychology

Four hundred children were selected at random from the fourth grade (A and B) in four schools in the Philadelphia district chosen by the Director of Educational Research of the system as giving a sampling of the city schools. Of the group, two hundred were boys and two hundred girls. That these numbers were equal was purely a matter of chance.

The method of examination and the treatment of the results followed that of Altmaier in his investigation of the sixth grade.

The statistical findings are given in detail in the tables found on pp. 34–40. The relative difficulty of the formboards is strikingly shown in the number of failures occurring on each. Thus there were no failures on either trial of the Witmer Formboard—which at this level merits a place in the battery of tests chiefly as a means of establishing rapport, of acquainting the subject with the nature of the behavior required in solving formboards and in arousing the competitive spirit in the child by giving him a feeling of efficiency. More difficult is the Witmer Cylinder test on which 11.75 per cent of the total group failed the first trial. Only one individual, however, failed on the second trial. And finally, on the first trial of the Dearborn Formboard 1C (Witmer) 62.00 per cent and on the second trial 30.75 per cent failed to solve the problem in the required time. As contrasted with the Witmer Formboard, this test represents for this group a real problem, demanding a concentration of abilities and taxing the competency of the individual to its limit.

It is interesting to note that the median scores on the three memory span tests are identical with those obtained on the fifth and sixth grade groups. The mean and median intelligence quotients differ from the expected score of 100 by less than half a point.

and for the training in the clinical field which have made an investigation of this nature possible and attractive; to Dr. Miles Murphy, Assistant Professor in Psychology at Pennsylvania, for the aid given in consultation on problems arising throughout the period of investigation, and to Dr. Carl L. Altmaier, at the present time Instructor in Psychology at Trinity College, Hartford, Conn., for assistance of a practical nature in the inauguration of the investigation. The author desires further to extend his thanks to Dr. Morris S. Viteles, Assistant Professor at Pennsylvania, for valuable suggestions on the statistical treatment of the results.

The examiners were Jay L. Otis, Wendell R. Carlson, Charles A. Limburg, and Austin G. Wells, Assistant Instructors in Psychology at Pennsylvania, Franklyn Fry, Lambertus Wartena, and Sadie Goldman, Graduate Students, and the author. At each session there were no more than four examiners, one of whom was always the author.

A sex difference on the formboards was noted by Leaming, Easby-Grave, Murphy and Altmaier. In the results obtained in the present study, statistically significant differences favoring the males were found between the means on the two trials of the Dearborn Formboard 1C (Witmer), and differences approximating statistical significance on both trials of the Witmer Cylinder test.[8] The percentages of failure on these tests are as follows:

	Males	Females
CYL 1st trial	8.00%	15.50%
DFB 1st trial	49.50	72.50
DFB 2nd trial	23.50	38.00

With the exception of the first trial of the Dearborn test, however, the degree of overlapping is exceedingly high as seen from the following:

CYL 1. 40.08 per cent of the females reach or exceed the median of the males, giving a difference of 9.92 per cent.

CYL 2. 46.85 per cent of the females reach or exceed the median of the males, giving a difference of 3.15 per cent.

DFB 1. 25.32 per cent of the females reach or exceed the median of the males, giving a difference of 24.68 per cent.

DFB 2. 40.55 per cent of the females reach or exceed the median of the males, giving a difference of 9.45 per cent.

On the Witmer Formboard, the memory span tests, and the intelligence quotients, no significant differences between the sexes is observed.

Part II

The Problem

Apart from the actual determination of the decile norms for the various groups, those conducting the earlier investigations have studied many characteristics of the groups concerned. Thus Easby-Grave analyzed the group with reference to the nationality of the father, discovering that over half of the fathers were born in this country, while Italy and Russia contributed thirty-three per cent, with the remainder divided among sixteen other countries. Alt-

[8] It should be noted that although as part of the determination of the significance of differences between groups the formula: difference between the means divided by the sigma of the difference, has been used, the skewness of the curves of the formboard scores makes the index so obtained of questionable significance. For this reason the percentages of failure and the degree of overlapping have been calculated.

maier's analysis according to occupation of the father showed that the group represented the average or modal family and that, following Taussig, it belonged to the class of skilled workmen and lower middle class. Both Easby-Grave and Murphy studied the Binet results with reference to the number of cases passing the individual tests. Murphy too made an analysis of the basal ages produced by the group and introduced the median modal age group. Altmaier compared this median modal age group with the total group and found no statistically significant differences between the two. Buzby,[9] analyzing the results of Murphy and McDermott-Murphy, found that the middle 40 per cent of the age range, and indeed the middle 20 per cent were no more homogeneous than the total group except with regard to the Intelligence Quotient. Altmaier made the further comparison of the performance results of the A and B sections of the grade and found no significant differences between these and the entire group. The A and B sections were not significantly different in either performance tests or in intelligence quotients, while the difference in mental age found is to be expected on the basis of increased chronological age.

We see, then, that much is known about the groups on which the Witmer Clinical Standards have been obtained, both with respect to background and internal variability. There is, however, one problem which has confronted the clinicians using the standards in practice, a problem which up to the present time it has been impossible to solve.

Those conducting the previous investigations have faced the problem of the selection of the proper group. A discussion of this difficulty is given by Murphy. The general procedure has been to take a random sampling of a school grade, the children being selected, as in the present study, from several sections of the city in order to obtain subjects from superior, average, and poor environments. This procedure has given adequate norms for the particular grades. The question has often arisen, however, as to the degree to which the norms for a grade coincide with the norms for the theoretically corresponding age level. Most of the investigators believed in a close agreement, as is shown by the titles of their studies which indicate that they deal with "The Six Year Old Performance Level," and the "Ten Year Level of Competency." There has been a tendency, too, in clinical practice to use the grade norms and to give, however, ratings with respect to age level.

[9] Buzby, Dallas. A Study of Test Results at the Third and Fifth Grade Levels. *Psychol. Clin.*, 1931, 20, 1–29.

Now it is true that since the age of entrance to school is six years, the first grade should represent the six year level; the second grade, the seven; the third grade, the eight; the fourth, the nine; the fifth, the ten; and the sixth, the eleven year level. The rates of promotion, however, are such that every grade, far from demonstrating homogeneity with respect to chronological age, represents great variation in age.

The difficulty, of course, could be avoided by abandoning the reference to age level, with respect to these norms and using them solely as grade levels of performance. For obvious reasons, however, there has developed in the field of mental testing, the tradition of using the age level for comparison of demonstrated behavior. In pre-school testing the chronological age can hardly be dispensed with. With adults, a grade level has little value. It has seemed, then, worth while to investigate the problem and to attempt to determine whether the grade level norms do represent an age level. Despite the variation in range of chronological age between a grade group and a normal group of that age theoretically corresponding to the grade, is the composition of the two groups with respect to the quantitative behavior similar to such an extent that for clinical purposes the norms of the group show no significant difference? The group representing one age level is homogeneous with respect to this factor but not in regard to grade, while the group representing one grade presents the opposite characteristics. The question, then is whether, despite these differences the two groups are for our purpose sufficiently homogeneous with reference to other, fundamental factors.

Altmaier made the first approach to this problem when he isolated each age group within his total group and compared the subgroups with the total. His findings were to the effect that "no one group . . . is sufficiently like the entire group in average score so that it can be said to be representative of the entire group." This does not, however, solve our problem since each age group within the total group represents with reference to the entire group of that particular age a very selected part, that is, that part of the total age group which has progressed to but not beyond the particular grade. It is not surprising, therefore, that the grade group, composed of individuals showing great variation with respect to age and competency, should display significant differences from any of the isolated, highly selected sub-groups.

With the completion, however, of the present investigation, the

necessary data for the determination of the relation of grade norms to those of the theoretically corresponding age group have for the first time been at hand. We now have the records of the performance of groups in the first, second, third, fourth, fifth, and sixth grades—that is, of all those grades in which are to be found children at the nine year chronological age level, the age theoretically corresponding to the fourth grade.

The comparison of the performance measures of the Fourth Grade children with those of the children at the Nine Year Chronological Age level accordingly constitutes the second part of this study.

Procedure

Since the groups of the previous studies were selected at random we may presume that the nine year old children in those groups also represent a random sampling of the nine year old children in the various grades. Defining the nine year old group as including those whose chronological ages range from 8 years, 11 months, and 16 days to 9 years, 11 months, and 15 days, those cases in the investigations of the first, second, third, fourth, fifth and sixth grades falling within this range were taken to represent a random sampling of nine year old children.

The cases in the first grade were taken from an investigation, more recent than that of Easby-Grave, conducted by Richards.[10] Those in the second grade were included in a study now being made by Wood.[11] The investigation of the third grade, conducted by McDermott-Murphy, supplied the children of that grade. Cases in the fifth grade were taken from the results of the study made by Murphy, while the investigation by Altmaier gave the cases of the sixth grade.

The distribution of these cases, numbering 406, is as follows:

Grade	1st	2nd	3rd	4th	5th	6th
Original Group (cases)	400	400	400	400	500	400
Nine Year Old (cases)	6	15	114	197	68	6
Per cent of Nine Year Total Group	1.4	3.6	28.0	48.5	16.7	1.4

[10] Richards, T. W. Psychological Tests in the First Grade. *Psychol. Clin.*, 1933, 21, 235–242.

[11] Wood, Sarah. Unpublished manuscript in the Department of Psychology, University of Pennsylvania.

In the studies made on the first and second grades, the Dearborn Formboard 1C (Witmer) was not administered because of the difficulty of the test for those levels. The data for the nine year group on the two trials of this test include, then, only the results for the third, fourth, fifth and sixth grades.

Eight groups were isolated for comparison: the Total Fourth Grade Group of 400 cases, the Fourth Grade Nine Year Group of 197 cases, the Total Nine Year Group of 406 cases, the Nine Year Group of Other Grades of 209 cases, the Fourth Grade Males of 200 cases, the Fourth Grade Females of 200 cases, the Nine Year Males of 193 cases, and the Nine Year Females of 213 cases. In treating the results for the Intelligence Quotients, another group was isolated, the Fourth Grade Group of Other Ages. By the Nine Year Group of Other Grades is meant those nine year old children not in the fourth grade, and by the Fourth Grade Group of Other Ages is meant those in the fourth grade who are not nine years of age.

Throughout this study the following abbreviations for the several groups are employed: Fourth Grade Total (4-T), Nine Year Total (9-T), Fourth Grade Males (4-M), Fourth Grade Females (4-F), Nine Year Males (9-M), Nine Year Females (9-F), Fourth Grade Nine Year Group (4-9 and 9-4), Nine Year Group of Other Grades (9-X), and Fourth Grade Group of Other Ages (4-Y).

As had already been done for the 4-T, 4-M and 4-F groups deciles were determined for the 9-T, 9-M and 9-F groups. The mean, sigma of the mean, sigma of the distribution, and coefficient of variation, together with the skewness for the various tests were determined for the several groups. By means of the formula previously mentioned the following comparisons were made:

4-T with 9-T.....	4-T with 9-X	9-M with 9-F
4-M with 9-M.....	4-9 with 9-X	
4-F with 9-F.....	4-T with 4-9	

In one point the statistical treatment differed from that previously indicated. Since the 4-9 group is a subgroup of 4-T, for the 4-T—4-9 comparison the following formula was used:[12]

$$\sigma \text{ diff.} = \sqrt{\frac{\Sigma^2}{N} + \frac{\sigma^2}{n}\left(1 - \frac{2n}{N}\right) - \frac{n\,(M-m)^2}{N(N-n)^2}}$$

[12] Pearson, K. Note on the Significant or Non-significant Character of a Sub-sample Drawn from a Sample. *Biometrika*, 1906–7, 5, 181–183.

in which

Σ = standard deviation of the total group,
N = number of cases of the total group,
M = average of the total group,
σ = standard deviation of the subgroup,
n = number of cases of the subgroup,
m = average of the subgroup.

Results

Performance Tests. No statistically significant differences were found between the means of any of the groups on either trial of the Witmer Formboard.

Although differences approaching statistical significance were discovered on the first trial of the Cylinders between the means of the 4-T and 4-9 groups and between the 4-9 and 9-X groups, no such difference occurred between the means of the 4-T and 9-T groups. Likewise on the second trial an approach to a significant difference is seen between the means of the 4-T and 4-9 groups but not between those of the 4-T and 9-T groups. The percentages of failures on the first trial are as follows: 4-T—11.75; 9-T—12.83; 4-9—13.70; 9-X—11.48.

The median of the 4-T group is equalled or surpassed by 51.9 per cent of the 9-T group, giving a difference of but 1.9 per cent.

On the Dearborn Formboard, likewise, no significant difference was found on either trial between the means of the 4-T and 9-T groups. No such difference appeared between the means of the 4-T and 4-9 groups although between the 4-T and 9-X and between the 4-9 and 9-X groups significant differences occurred. The percentages of failures for the groups are as follows:

	4-T	9-T	4-9	9-X
1st trial	62.00	57.40	62.43	51.99
2nd trial	30.75	25.45	28.42	22.34

The difference in the percentage of failure for the 4-T and 9-T groups is therefore but 4.60 per cent on the first trial and 5.30 per cent on the second trial.

Calculating the percentile ratings which the scores delimiting the deciles of the 4-T group obtain when compared with the 9-T distribution we get the following results:

PERCENTILES: 1ST TRIAL

Score	4-T	9-T	Difference
144	90	87.9	2.1
202	80	76.2	3.8
256	70	64.6	5.4
			11.3 Average—3.8

PERCENTILES: 2ND TRIAL

Score	4-T	9-T	Difference
67	90	90.9	0.9
89	80	79.5	0.5
111	70	69.2	0.8
134	60	58.4	1.6
175	50	46.0	4.0
222	40	33.3	6.7
			14.5 Average—2.4

The similarity of the two total groups is also seen in the percentages of failures on the formboards according to sex as shown in the table:

		4	9
CYL 1st trial	Male.....	8.00	9.32
	Female.....	15.50	15.49
DFB 1st trial	Male.....	49.50	46.48
	Female.....	72.50	65.85
DFB 2nd trial	Male.....	23.50	18.37
	Female.....	38.00	31.31

The difference between the means of the 9-M and 9-F groups are, as was seen with the 4-M and 4-F groups, statistically significant.

Memory Span. Inasmuch as the results of an investigation of this type are used entirely in the form of decile tables, by means of which a given performance is placed with reference to those superior and those inferior, any difference between the groups to be significant for clinical purposes must be manifest in these deciles. A glance at the decile tables for the memory span results shows the obvious lack of need for treatment of the scores for the comparison of the groups. For all groups the deciles for the Audito Vocal Digit Forward and the Visual Audito Vocal Digit Forward

Memory Span tests are identical, while in the Audito Vocal Digit Reverse Memory Span test the only difference lies in the fourth decile for the 9-T and 9-M groups.

Intelligence Quotient. In the Intelligence Quotients we do find a difference between the means of the 4-T and 9-T groups of 3.00 which is statistically significant. An analysis of the 4-T and 9-T groups gives the following means:

	4th Grade	Nine Year		Nine Year Group by Grades		
4-9	103.99	103.99	9-4	6th Grade	121.8	P.E. 5.23
4-T	99.97	102.97	9-T	5th	114.3	1.04
4-Y	96.74	101.81	9-X	4th	103.9	.59
				3rd	97.4	.76
				2nd	84.2	1.98
				1st	69.1	1.76

The greater difference between the means of the 4-Y and 9-X groups is also statistically significant. The 4-Y group has the lowest mean of the five groups and all the nine year groups are superior to the 4-T group, not, however, in the case of the 9-X group, sufficiently superior to be significant.

Since a comparison of the curves and decile points of the 4-T and 9-T groups gave the impresssion that the essential feature was similarity and not difference, a more thorough investigation into the value of the difference between means as a criterion of similarity and dissimilarity was made.

It was found that although the difference between the medians of the 4-Y and 9-X groups (5.8) is significant (z equal to 2.82), the difference between the medians of the 4-T and 9-T groups was 2.5, a difference not statistically significant (z equal to 1.89).

The calculation of the percentage of overlap revealed that 57.2 per cent of the 9-T group reached or exceeded the median of the 4-T group, giving a difference of but 7.2 per cent while 43.75 per cent of the 4-T group reached or exceeded the median of the 9-T group, giving a difference of 6.25 per cent. Calculating these differences for each decile revealed an average difference of but 5.4 per cent when the 4-T group is compared with the decile points of the 9-T group and of only 5.53 per cent when we compare the 9-T group with the decile points of the 4-T group.

On the first trial of the Witmer Formboard, the 9-T group is almost twice as variable as the 4-T, and on the second trial some-

what more variable. On the Witmer Cylinders, the Dearborn Formboard 1C (Witmer), and the Intelligence Quotients, the difference in variability (as determined by the coefficient of variation) is so small as to be insignificant. The 4-9 group tends to be the least variable of the 9-X group to be the most variable. In the Intelligence Quotients, however, the 4-Y group shows a slightly greater variation than the 9-X group.

Discussion

In so far as our statistical procedure is valid in determining the significance of differences no such difference appears between the total groups on the formboards or the memory span tests. Where the number of failures is great, the percentages of failures show little difference and the degree of overlapping is extremely high, with the differences in percentages for the median score and for those delimiting the deciles being themselves less than a decile.

Again, although the difference between the means on the Intelligence Quotients is statistically significant, the difference between the medians is not significant and the degree of overlapping is again so large as to lessen materially the value of the difference between the means as an index of similarity.

However, as noted before, the norms derived from these investigations—the Witmer Clinical Standards—are employed entirely in the form of decile tables, and our problem has been the determination of the validity of using the decile norms obtained on a grade group and at the same time interpreting the ratings in terms of the theoretically corresponding age group.

The final decision on this question must depend upon a comparison of the ratings obtained for the range of scores when compared with the two total groups. This was accomplished by determining those scores on all the tests which gave different decile ratings when compared with the norms of the two groups, and ascertaining the number of cases of the 9-T group which fall at these scores. From this we ascertain the percentage of the group the scores of which have the same rating when compared with the deciles [13] of the fourth grade total group as when compared with the nine year total group. These figures are as follows.

[13] For scores delimiting deciles correspondence within the quintile rating was necessarily used in this calculation.

	Same	Different		Same	Different
WFB 1.....	97.54%	2.46%	CYL 1.....	93.10%	6.90%
WFB 2.....	100.00%	0.00%	CYL 2.....	95.08%	4.92%
DFB 1.....	86.91%	13.09%	IQ.....	56.10%	43.90%
DFB 2.....	89.16%	10.84%			

All "different" ratings were a decile off and, with negligible exceptions, were in the same direction: the 4-T ratings, as compared with the 9-T ratings, were for the Cylinders [14] too low and for the Dearborn [15] test too high.

If we compare the performance of a nine year old child with the norms of the fourth grade group, we have, therefore, almost perfect assurance that for the Witmer Formboard the rating so obtained is the same as that from the age norms, while for the Intelligence Quotient the chances are about even that the rating is the same and that it is a decile too high. For the Witmer Cylinders and the Dearborn test there are nine chances in ten of the same rating with both norms and one chance in ten that the rating will be different—a decile too low for the Witmer Cylinders, a decile too high for the Dearborn. The ratings for the Memory Span are, of course, identical for both norms.

In relating these results to our original question concerning the interpretation in terms of an age level of ratings obtained from grade norms, several points must be considered.

An even chance that the decile rating from grade norms is the same as that to be obtained from the age norms, or even nine chances in ten of the same rating, would be unsatisfactory for clinical use were the extent and direction of the possible error not known. With the level investigated here this data has been presented, so that we can be sure that a particular quintile rating obtained from the grade norms will include the rating to be obtained from the age norms. And while a decile rating is certainly more valuable in practice, a quintile rating is at times of some use.

However, even if we assume the chances of obtaining identical ratings from the grade and age norms and the extent of the possible error as found in this study to be the same for all age-grade levels, there is no reason to believe that the direction of error will

[14] On the second trial one score, at which one individual (.02 per cent of the group) placed, would receive a rating a decile too high.

[15] On the second trial one score, at which three individuals (.73 per cent of the total group) placed, would receive a rating a decile too low.

necessarily be the same as that found with this group. It follows that even with the chances of identity and the extent of possible error constant for all the age-grade groups, the finest rating from grade norms which we could be certain included the rating to be obtained from the corresponding age norms would be in a group of thirty per cent—composed of the decile containing the score produced and the two adjoining deciles. But since thirty per cent of the scale of competency may include both definitely normal and feebleminded individuals, it is obvious that such a rating is entirely too coarse to be of significance in arriving at a diagnosis.

The final decision to our problem must then be that although our grade and age groups closely approach each other in performance, the dissimilarity is, nevertheless, of sufficient magnitude to render the interpretation of ratings obtained from the norms of one group in terms of relative superiority in the other of little clinical value.

Conclusions

1. Statistically no significant difference exists between the results of the fourth grade group and those of the nine year old group.

2. In the actual use of the norms, however, the translation of ratings obtained from the grade results into ratings on the age scale will give classifications on the latter so coarse as to be of extremely limited value in clinical practice.

3. Only on the first trial of the Dearborn Farmboard 1C (Witmer) is the evidence for a significant sex difference consistent. This difference is in favor of the males.

Bibliography

Altmaier, Carl L. The Performance Level of Children in the Sixth Grade in Two Philadelphia Public Schools. *Psychol. Clin.*, 1931, 19, 233–257.

Buzby, Dallas E. A Study of Test Results at the Third and Fifth Grade Levels. *Psychol. Clin.*, 1931, 20, 1–29.

Easby-Grave, Charlotte. Tests and Norms at the Six Year Old Performance Level. *Psychol. Clin.*, 1924, 15, 261–300.

Hallowell, Dorothy. Mental Tests for Pre-School Children. *Psychol. Clin.*, 1928, 16, 233–276.

Leaming, Rebecca E. Tests and Norms for Vocational Guidance at the Fifteen Year Old Performance Level. *Psychol. Clin.*, 1922, 14, 193–220.

Murphy, Miles. The Ten Year Level of Competency. *Psychol. Clin.*, 1928, 17, 33–60.

Richards, T. W. Psychological Tests in The First Grade. *Psychol. Clin.*, 1933, 21, 235–242.

Witmer, Lightner. Psychological Diagnosis and the Psychonomic Orientation in Analytical Science. *Psychol. Clin.*, 1925, 16, 1–18.

TABLE I

WITMER FORMBOARD

First Trial

Groups	Mean	σ mean	Sigma	Coef. Var.	Skewness
4-T	31.25	.47	9.44	30.21	.49
4-M	30.00	.48	6.85	22.83	.32
4-F	32.50	.80	11.30	34.76	.61
9-T	32.50	.90	18.15	55.85	.49
9-M	31.59	.61	8.42	26.65	.49
9-F	34.60	1.65	24.15	69.79	.52
9-4	31.60	.83	11.70	37.02	.72
9-X	34.00	1.62	23.40	68.82	.74

Second Trial

4-T	26.01	.36	7.20	27.68	.62
4-M	26.10	.53	7.60	29.12	.59
4-F	25.93	.48	6.85	26.42	.62
9-T	26.98	.50	10.10	37.43	.67
9-M	26.81	.59	7.85	29.24	.34
9-F	27.13	.81	11.80	43.49	.69
9-4	26.35	.54	7.70	29.22	.69
9-X	27.58	.85	12.00	43.51	.67

Comparison of Groups

Groups	First Trial			Second Trial		
	Diff.	σ diff.	Z	Diff.	σ diff.	Z
4-T –9-T	1.25 9	1.01	1.23	.97 9	.61	1.59
4-T –4-9	.35 9	.48	.73	.34 9	.41	.81
4-T –9-X	2.75 9	1.68	1.64	1.57 9	.92	1.70
4-9 –9-X	2.40 4	1.82	1.31	1.23 X	1.00	1.23
4-M–4-F	2.50 F	.93	2.68	.17 M	.71	.23
9-M–9-F	3.01 F	1.75	1.72	.32 F	1.11	.45
4-M–9-M	1.59 9	.76	2.09	.71 9	.79	.89
4-F –9-F	2.10 4	1.83	1.14	1.20 9	.93	1.29

Note: The numbers and letters immediately following the differences indicate the groups having the higher mean. The *Z* column represents the quotient obtained by dividing the Difference by its sigma. Unless otherwise indicated the coefficients of skewness are positive.

TABLE II

WITMER CYLINDERS

First Trial

Group	Mean	σ mean	Sigma	Coef. Var.	Skewness
4-T	147.06	4.08	81.60	55.48	.95
4-M	136.15	5.42	76.50	56.18	.90
4-F	157.98	6.01	84.75	53.64	.82
9-T	149.65	4.05	81.75	54.60	.93
9-M	142.45	5.42	75.45	52.96	.79
9-F	156.25	5.92	86.55	55.39	1.00
9-4	157.51	5.77	81.00	51.42	.95
9-X	142.30	5.62	81.00	56.52	.99
Second Trial					
4-T	83.80	2.15	42.90	51.19	.81
4-M	77.50	2.63	37.20	48.00	.75
4-F	90.03	3.32	46.80	51.88	.81
9-T	86.20	2.24	45.30	52.55	.86
9-M	79.15	2.74	38.10	48.13	.83
9-F	92.50	3.42	49.95	54.00	.78
9-4	89.20	3.02	42.45	47.58	.97
9-X	83.20	3.29	47.40	56.97	.77

Comparison of Groups

Groups	First Trial			Second Trial		
	Diff.	σ diff.	Z	Diff.	σ diff.	Z
4-T –9-T	2.59 9	5.66	.45	2.40 9	3.10	.77
4-T –4-9	10.45 9	4.14	2.53	5.40 9	2.17	2.48
4-T –9-X	4.76 4	6.95	.68	.60 4	3.92	.15
4-9 –9-X	15.20 4	7.97	1.90	6.00 4	4.45	1.34
4-M–4-F	21.83 F	8.09	2.69	12.53 F	4.23	2.96
9-M–9-F	13.80 F	8.02	1.72	13.35 F	4.38	3.02
4-M–9-M	6.30 9	7.65	.82	1.65 9	3.79	.45
4-F –9-F	1.73 4	8.43	.20	2.48 9	4.75	.52

TABLE III

DEARBORN FORMBOARD 1C (WITMER)

First Trial

Group	Mean	σ mean	Sigma	Coef. Var.	Skewness
4-T	264.85	3.40	68.05	25.69	−1.71
4-M	248.13	5.35	75.45	30.41	−2.04
4-F	281.65	3.86	54.45	19.33	−1.33
9-T	257.95	3.60	70.65	27.38	−1.91
9-M	247.15	5.47	73.35	29.67	−1.75
9-F	267.40	4.68	67.05	25.07	−1.66
9-4	268.15	4.57	64.20	23.94	−1.68
9-X	247.30	5.48	75.15	30.38	−2.16

Second Trial

4-T	188.88	4.87	97.35	51.54	.45
4-M	174.18	6.76	95.40	54.77	.62
4-F	203.58	6.89	97.20	47.74	.002
9-T	180.70	4.71	92.40	51.13	.59
9-M	169.50	6.74	90.30	53.27	1.11
9-F	190.45	6.50	93.00	48.83	.54
9-4	188.35	6.70	94.05	49.93	.52
9-X	173.35	6.52	89.40	51.57	.76

Comparison of Groups

Groups	First Trial			Second Trial		
	Diff.	σ diff.	Z	Diff.	σ diff.	Z
4-T –9-T	6.90 4	4.95	1.39	8.18 4	6.77	1.21
4-T –4-9	3.30 4	3.44	.96	.53 4	4.92	.12
4-T –9-X	17.55 4	6.45	2.72	15.33 4	8.13	1.91
4-9 –9-X	20.85 4	7.13	2.92	15.00 4	9.34	1.61
4-M–4-F	33.53 F	6.59	5.08	29.40 F	9.65	3.04
9-M–9-F	20.25 F	7.19	2.81	20.96 F	9.36	2.23
4-M–9-M	.98 4	7.65	.12	4.68 4	9.54	.49
4-F –9-F	14.24 4	6.16	2.35	13.13 4	9.47	1.38

TABLE IV

INTELLIGENCE QUOTIENT

Group	Mean	σ mean	Sigma	Coef. Var.	Skewness
4-T	99.97	.77	15.45	15.47	.07
4-M	97.53	1.10	15.52	15.91	−.28
4-F	100.43	1.05	14.92	14.83	.04
9-T	102.97	.73	14.81	14.38	.17
9-M	103.17	1.11	14.60	14.15	.28
9-F	102.78	1.02	14.95	13.57	.02
9-4	103.99	.88	12.35	11.87	.14
9-X	101.81	1.14	16.55	16.25	.14
4-Y	96.74	1.18	16.85	17.41	.27

Comparison of Groups

Group	Diff.	σ diff.	Z
4-T –9-T	3.00 9	1.06	2.83
4-T –4-9	4.02 9	.78	5.15
4-T –9-X	1.84 X	1.38	1.33
4-9 –9-X	2.18 4	1.44	1.51
4-M–4-F	2.90 F	1.52	1.90
9-M–9-F	.39 M	1.51	.26
4-M–9-M	5.64 9	1.56	3.62
4-F –9-F	2.35 9	1.77	1.32
4-Y –9-T	6.13 9	1.38	4.51
4-Y –9-4	7.25 9	1.47	4.93
4-Y –9-X	5.07 9	1.64	3.09

Decile Tables

Nine Year Total

C.A.	M.A.	%	I.Q.	A.V.	V.A.V.	Rev.	WFB1	WFB2	CYL1	CYL2	DFB1	DFB2
9–11	15–4	100	160.4	9	10	6	15	12	24	29	59	25
9–11	13–7	99	143.7	8	10	6	16	15	44	37	71	42
9–10	11–6	90	121.6	7	8	5	22	19	65	49	129	69
9–9	10–10	80	114.7	6	7	4	25	21	80	55	185	88
9–8	10–5	70	109.2	6	7	4	27	22	93	62	236	110
9–7	10–0	60	105.4	6	7	4	29	23	105	67	288	131
9–6	9–7	50	102.1	6	7	4	30	25	124	73	F	162
9–5	9–3	40	98.5	5	6	4	33	26	144	81	F	195
9–3	9–0	30	94.7	5	6	3	35	28	174	91	F	252
9–2	8–9	20	91.3	5	6	3	38	30	223	107	F	F
9–1	8–3	10	85.6	5	5	3	45	35	F	134	F	F
9–0	7–0	1	68.9	4	4	0	86	75	F	F	F	F
9–0	5–6	0	59.4	4	3	0	285	100	F	F	F	F

Nine Year Males

C.A.	M.A.	%	I.Q.	A.V.	V.A.V.	Rev.	WFB1	WFB2	CYL1	CYL2	DFB1	DFB2
9–11	15–4	100	160.4	9	10	6	15	15	24	29	70	25
9–11	13–7	99	143.7	8	9	6	16	15	44	37	78	35
9–10	11–7	90	123.5	7	8	5	23	19	65	47	131	64
9–9	10–9	80	114.3	6	7	4	26	21	81	53	161	84
9–8	10–4	70	108.6	6	7	4	27	22	92	58	202	99
9–7	10–0	60	104.7	6	7	4	29	24	102	63	245	120
9–6	9–7	50	101.8	6	7	4	30	26	123	69	294	151
9–4	9–3	40	98.2	5	6	3	32	27	137	74	F	182
9–3	9–0	30	94.8	5	6	3	34	29	155	83	F	219
9–2	8–9	20	93.3	5	6	3	37	31	202	100	F	296
9–1	8–3	10	86.5	5	5	3	42	36	281	129	F	F
9–0	7–0	1	77.7	4	4	2	55	51	F	205	F	F
9–0	5–6	0	60.5	4	3	0	86	59	F	F	F	F

Nine Year Females

C.A.	M.A.	%	I.Q.	A.V.	V.A.V.	Rev.	WFB1	WFB2	CYL1	CYL2	DFB1	DFB2
9–11	14–4	100	151.3	9	10	6	16	12	42	35	59	40
9–11	13–11	99	145.9	7	10	6	16	14	43	37	59	42
9–10	11–6	90	120.6	7	8	5	21	18	65	51	154	73
9–9	10–11	80	115.0	6	8	4	25	21	79	59	208	96
9–8	10–5	70	109.8	6	7	4	27	22	94	66	260	123
9–7	10–1	60	106.3	6	7	4	29	23	109	72	F	141
9–6	9–9	50	102.7	6	7	4	30	24	127	79	F	174
9–5	9–3	40	99.2	5	6	4	33	26	152	88	F	210
9–4	9–0	30	94.7	5	6	3	35	28	184	98	F	F
9–2	8–9	20	91.5	5	6	3	40	30	262	110	F	F
9–1	8–0	10	84.3	5	5	3	47	34	F	144	F	F
9–0	6–3	1	68.8	4	4	0	122	80	F	F	F	F
9–0	5–6	0	59.4	4	3	0	285	100	F	F	F	F

Fourth Grade Total

C.A.	M.A.	%	I.Q.	A.V.	V.A.V.	Rev.	WFB1	WFB2	CYL1	CYL2	DFB1	DFB2
15–5	13–7	100	141.5	9	10	7	15	13	24	30	54	23
13–3	13–3	99	139.3	8	9	6	16	16	40	37	67	31
11–6	11–3	90	121.0	7	8	5	22	19	61	47	144	67
10–7	10–8	80	113.3	6	7	4	25	21	77	54	202	89
10–1	10–3	70	108.1	6	7	4	26	22	92	60	256	111
9–10	10–0	60	103.6	6	7	4	28	23	105	66	F	134
9–8	9–8	50	99.6	6	7	4	30	25	121	72	F	175
9–6	9–6	40	95.2	5	6	4	32	26	142	80	F	222
9–4	9–1	30	91.8	5	6	3	34	28	172	90	F	F
9–1	9–0	20	87.1	5	6	3	37	30	216	103	F	F
8–10	8–9	10	82.1	5	5	3	42	34	F	136	F	F
8–3	7–9	1	65.3	4	4	3	67	58	F	278	F	F
7–11	7–3	0	50.2	4	3	2	122	61	F	F	F	F

Fourth Grade Males

C.A.	M.A.	%	I.Q.	A.V.	V.A.V.	Rev.	WFB1	WFB2	CYL1	CYL2	DFB1	DFB2
14–2	13–7	100	141.5	9	9	6	15	13	24	35	54	23
13–6	13–4	99	139.3	9	9	6	16	15	39	35	61	25
11–7	11–4	90	122.8	7	8	5	22	18	59	45	126	56
11–0	10–8	80	112.2	6	7	4	24	20	72	51	164	83
10–5	10–4	70	107.2	6	7	4	26	22	85	56	206	105
9–11	10–1	60	103.0	6	7	4	28	23	99	62	256	120
9–9	9–9	50	99.0	6	7	4	29	25	113	68	300	154
9–7	9–6	40	94.7	5	6	4	31	26	129	74	F	192
9–4	9–3	30	91.4	5	6	3	33	28	153	82	F	242
9–2	9–0	20	86.5	5	6	3	35	30	201	95	F	F
8–10	8–6	10	80.4	5	5	3	39	34	278	123	F	F
8–2	7–9	1	65.3	4	4	3	46	52	F	205	F	F
7–11	7–3	0	64.4	4	3	2	58	59	F	278	F	F

Fourth Grade Females

C.A.	M.A.	%	I.Q.	A.V.	V.A.V.	Rev.	WFB1	WFB2	CYL1	CYL2	DFB1	DFB2
15–5	13–3	100	141.1	8	10	7	16	16	35	30	71	26
12–9	12–6	99	136.3	8	9	6	16	16	40	39	80	35
10–10	11–2	90	119.7	7	8	5	22	20	64	50	196	77
10–4	10–8	80	114.4	6	7	4	25	21	83	58	256	95
10–0	10–2	70	109.0	6	7	4	27	22	97	65	F	122
9–9	9–11	60	104.2	6	7	4	29	23	113	71	F	151
9–7	9–7	50	100.2	6	7	4	30	25	135	77	F	204
9–5	9–3	40	95.9	5	6	4	32	26	157	87	F	286
9–2	9–1	30	92.2	5	6	3	34	28	187	97	F	F
9–0	9–0	20	87.7	5	6	3	38	29	256	109	F	F
8–10	8–6	10	81.8	5	5	3	44	33	F	144	F	F
8–4	7–9	1	68.2	4	5	3	71	59	F	297	F	F
8–0	7–3	0	50.2	4	4	2	122	61	F	F	F	F

LANCASTER PRESS, INC., LANCASTER, PA.